What is continuous optimization?

Classical optimization
$$\boxed{\min_{x \in X} f(x) \qquad (*)}$$

on $X \subseteq \mathbb{R}^n$ with possibly $n$ large. Alternatively $X = \mathbb{Z}^n$ being discrete optimization.

(*) allows to apply results like compactness arguments, i.e. $X \subset\subset \mathbb{R}^n$, $f$ cont $\Rightarrow$ existence of min.

or projection and separation arguments

$$X = \{ x \in \mathbb{R}^n : h(x) = 0 \}, \quad h: \mathbb{R}^n \to \mathbb{R}^m$$

If $\nabla h(x^*)$ l. indep., then $x^*$ min. implies $\exists! \lambda \in \mathbb{R}^m$

$$\nabla f(x^*) + \nabla h(x^*) \lambda = 0.$$

Here, $\nabla h = (\nabla h)_{ij} = \frac{\partial h_j}{\partial x_i}$ and $\lambda \in \mathbb{R}^m$.

Motivation: $x^*$ local min $\Rightarrow$ $f(x^*+y) \geq f(x^*)$ $\forall y \in \mathbb{R}^n$ s.t. $h(x^*+y) = 0$.

Taylor expansion gives
$$\begin{cases} \nabla f(x^*) y \geq 0 + o(\|y\|^2) \\ Dh(x^*) y = 0 + o(\|y\|^2) \end{cases}$$

For $x_y$ we obtain $Df(x^*)y = 0$ and

$Dh(x^*)y = 0$. Hence, $\nabla f(x^*) = Df(x^*)^T$ is orthogonal

to $y$ and $y \in \ker(Dh(x^*))$. In finite dimensions

we have

$$(\ker A)^\perp = \text{range}(A^T)$$

for all matrices $A \in \mathbb{R}^{n \times n}$.

Hence, $\nabla f(x^*) \in \text{range}(\nabla h(x^*)) \Rightarrow \nabla f(x^*) = \nabla h(x^*)\lambda$

for some $\lambda \in \mathbb{R}^m$. $\lambda$ is unique since

$$0 = \nabla h(x^*)(\lambda_1 - \lambda_2) \Rightarrow \lambda_1 = \lambda_2 \text{ since}$$

$\text{rg}(\nabla h(x^*)) = m$.

Neither compactness nor the relation for operators nor differentiability is given in the case where $X$ is infinite dimensional.

Examples of pb's where $X$ are infinite dimensional are control of ODEs, PDEs, shape optimization.

Typical pb's for those ~~are~~ are the following

Mayer Pb.

$$\min_{(x,u)} \varphi(x(T))$$

$$\text{s.t.} \quad \dot{x}(s) = f(x(s), u(s)) \quad , \quad x(0) = x_0$$

Robotics trajectory planning with $x$ being trajectory, $u$ being control. Also: active agents: $x = (x_1, \ldots, x_N)$

$f = (f_i) \quad f_i = \frac{1}{N} \sum_{j=1}^{N} (x_j - x_i) + u$ as opinion formation model.

PDE Pb.

$$\min_{(y,u)} \int_0^T \int_\Omega (y(s,x) - y_d)^2 + \frac{\alpha}{2} u(s,x)^2 \, ds\, dx$$

$$\text{s.t.} \begin{cases} y_t - \Delta y = u & \text{in } \Omega \subset \mathbb{R}^n \\ y = 0 & \text{on } \partial\Omega \end{cases}$$

Temperature tracking, fluid flow.

Distributed control $u = u(s,x)$ and state $y(s,x)$.

Shape Pb.

$$\min_{(\Omega_1, \Omega_2)} \int_\Omega g(y(x))\, dx$$

$$\text{s.t.} \quad -\nabla \cdot (a(x) \nabla y(x)) = 0 \quad \text{in } \Omega$$

$$y = 0 \quad \text{on } \partial\Omega$$

and $a(x) = \begin{cases} a_1 & x \in \Omega_1 \\ a_2 & x \in \Omega_2 \end{cases}$

and $\Omega_1 \cup \Omega_2 = \Omega$.

Heuristic approach to first two pb$^s$:

Finite dimensional case: $\min f(x)$ s.t. $h(x) = 0$,

introduce the Lagrange fct $L(x,\lambda) = f(x) + h(x)^T \lambda$

Then, if the Lagrange fct has a saddle pt $L(x^*,\lambda) \leq L(x^*,\lambda^*) \leq L(x,\lambda^*)$

then $x^*$ is local min. The saddle pt cond $L(x^*,\lambda^*) \leq L(x,\lambda^*)$ yields

$\left(0 = \frac{\partial L}{\partial x}\right)$ $\nabla f(x^*) + \nabla h(x^*)\lambda^* = 0$

and $L(x^*,\lambda) \leq L(x^*,\lambda^*)$ yields

$\left(0 = \frac{\partial L}{\partial \lambda}\right)$ $h(x^*) = 0.$

Consider the first pb. A possible Lagrange fct is

$$L(x,\lambda) = \varphi(x(T)) + \int_0^T \lambda(s)\left(\dot{x}(s) - f(x(s),u(s))\right) ds$$

for $L : C^1 \times C^1 \to \mathbb{R}$. Differentiation in fct space is formally similar to $\mathbb{R}^n$. Consider

$$J(x) := \int_0^T j(x(s))\, ds$$

Then, $\frac{dJ(x)}{dx}[h] = \lim_{t \to 0} \frac{J(x+th) - J(x)}{t}$ is the Gateaux derivative of $J$ if it exists for fixed $h$.

~~If $\mathcal{J}$ Frechet diff w/o t and~~

~~If $\mathcal{J}$ ... and~~ $\frac{d\mathcal{J}}{dx}(x)$ is a linear and

bounded operator, then it is Frechet diff.

Computation. $\frac{d\mathcal{J}}{dx}(x)[h] = \int_0^T \frac{1}{t}\left(j(x(s)+t\,h(s)) - j(x(s))\right) ds$

$$= \int_0^T \frac{1}{t}\left[t\, j'(x(s))h(s)\right] ds = \int_0^T j'(x(s))\, h(s)\, ds.$$

Note that $\mathcal{J} : C^0 \to \mathbb{R}$ and $\frac{d\mathcal{J}}{dx} \in L(C^0, \mathbb{R})$ and

since this space is the dual to $C^0$, i.e. the

space of Radon measures. Hence, $\frac{d\mathcal{J}}{dx}(x) = j'(x(s))ds$

and $\frac{d\mathcal{J}}{dx}[h] = \langle j'(x(s))ds, h \rangle_{M, C^0} = \int_0^T j'(x(s))h(s)\, ds.$

Differentiating the Lagrange fct we obtain

$$\frac{\partial}{\partial \lambda} L(x)[h_\lambda] = \int_0^T h_\lambda \left(\dot{x} - f(x,u)\right) ds \overset{!}{=} 0 \;\forall h_\lambda$$

$$\Rightarrow \quad \dot{x} = f(x,u)$$

$$L(x,\lambda) = \Psi(x(T)) + \int_0^T \lambda \dot{x} - \lambda f(x,u)\, ds$$

$$L(x,\lambda) = \Psi(x(T)) - \int_0^T \dot{\lambda} x - \lambda f(x,u)\, ds + (\lambda x)(T) - \lambda(0)x_0$$

$$\Rightarrow \frac{\partial L}{\partial x}(x,\lambda)[h_x] = \nabla\Psi(x(T))\, h_x(T) + \int_0^T -\dot{\lambda}\, h_x - \lambda f_x(x,u) h_x\, ds$$

$$+ \lambda(T)\, h_x(T) \quad = 0$$

Taking now a variation $h_x$ st. $h_x(T) \neq 0$ but $h_x = 0$ ae in $(0,T)$ we obtain

$$\nabla_x \varphi(x(T)) + \lambda(T) = 0.$$

and for the opposite case:

$$-\dot{\lambda}(s) = \lambda(s) f_x(x(s), u(s))$$

Taking a variation in $u$ yields

$$0 = \lambda(s) f_u(x(s), u(s))$$

The system is known as PMP. It comprises of a forward and backward eq for $x$ and $\lambda$ as well as an eq for $u$. Numerically, one possibility to solve the eq$^s$ are as follows:

Pick $u^k$, Solve $\dot{x}^k = f(x^k, u)$, $x^k(0) = x_0$ for $x^k$, Solve

$$-\dot{\lambda}^k = \lambda^k \, D f_x(x^k, u), \quad \lambda^k(T) = -\nabla \varphi(x^k(T)),$$

Update $u^k = u^{k-1} - \lambda^k(s) D f_u(x^k(s), u^k(s))$

This scheme is independent of the numerical method for the ODEs.

Why does it work? The pb $\min_{(x,u)} \varphi(x(T))$ sbj $\dot{x} = f(x,u)$, $x(0) = x_0$ is an unconstrained pb for $u$. Ie. for any given $u$ we can compute $x = x(u)$ through the constraints.

Therefore, we could also write

$$\min_u \Psi(x(T;u))$$

and solve this unconstrained pb by a gradient descent, i.e., $u^k = u^k - \frac{\partial}{\partial u}\Psi(x(T;u^k))$.

This requires to compute the gradient of $\Psi$ wrt to $u$.

$$\frac{\partial}{\partial u}\Psi(x(T;u))^{h_u(T)} = D\Psi(x(T;u))\,\frac{\partial x(T;u)}{\partial u}\,h_u(T)$$ for a variation at time $T$.

~~Computing $\frac{\partial x(T;u)}{\partial u}$ can be done by rewriting the ODE as~~ $\cancel{x(t) = x_0 + \int_0^t f(x(s),u(s))\,ds}$ ~~and~~

$$\cancel{\frac{\partial x(T)}{\partial u}(h) = \int_0^T \left(D f_x(x(s),u(s))\frac{\partial x(s)}{\partial u} + D f_u(\dots)\right) h_u(s)\,ds}$$

$$\cancel{\frac{\partial x(T)}{\partial u}[h_u(T)] = \int_0^T h_u(s)\left[D f_x \frac{\partial x(s)}{\partial u}\right] + h_u(s)\,D f_u\,h\,ds}$$

~~Hence, if $h_u(s) = \lambda(s)$ with $-\dot\lambda(s) = \lambda(s) D f_x$ and $0 = \lambda D f_u$~~

we obtain

$$\frac{\partial x(T)}{\partial u}\underbrace{h_u(T)}_{\lambda(T)} = \cancel{\int_0^T -\dot\lambda(s)\frac{\partial x(s)}{\partial u}\,ds}$$

Consider $(-\dot\lambda = \lambda D_x f) h_u(s)$ and $\dot x = f(x,u)$. Then

$$\frac{\partial}{\partial t}\left(\frac{\partial x}{\partial u}\right) = D_x f\,\frac{\partial x}{\partial u} + D_u f. \quad \| \text{ multiply by } h_u(s)$$

and hence

$$\left(\lambda\frac{\partial x}{\partial u}\right)' = -\lambda D_x f\,\frac{\partial x}{\partial u} + \lambda\left(D_x f\,\frac{\partial x}{\partial u} + D_u f\right) = \lambda D_u f \overset{\lambda D_u f = 0}{=} 0.$$

$$\Rightarrow \left(\lambda\frac{\partial x}{\partial u}\right)(s) = \left(\lambda\frac{\partial x}{\partial u}\right)(T)$$

For $\lambda(T) = D\Psi(x(T;u))$ we therefore get

Werkzeugmaschinenlabor WZL
der RWTH Aachen
www.wzl.rwth-aachen.de
info@wzl.rwth-aachen.de

$\lambda \frac{\partial x}{\partial u} = \text{cst}$

$$\frac{\partial}{\partial u}\psi = D_x \psi \frac{\partial x}{\partial u} h_u = \lambda(T) \frac{\partial x}{\partial u}(T)\, h_u(T) = \lambda(s) \frac{\partial x(s)}{\partial u}\, h_u(T)$$

$$= \int_0^s \lambda(s) D_x f(x(s), u(s))\, ds \; h_u(T)$$

~~If we apply a variation at time~~

~~Since~~

$\Rightarrow$ A representation of $\frac{\partial \psi}{\partial u}$ as a fct of $u$ for any $s$ is given by $\lambda(s) D_x f(x(s), u(s))$.

---

## Second pb (simplified)

$$\min \int_\Omega \frac{1}{2}|y - y_d|^2 \, dx + \frac{\alpha}{2}\int u^2 dx \quad \text{s.t.} \begin{cases} -\Delta y = u & \text{in } \Omega \\ y = 0 & \text{on } \partial\Omega \end{cases}$$

Constraint qualification $\nabla h(x)$ regular implies solvability of linearized pb. PDE is linear and expect the same eq. However, the sol theory usually requires weak form. Hence, the lagrangian is

$$L(y, u, p) = \int_\Omega \frac{1}{2}(y - y_d)^2 dx + \int -\nabla p \nabla y - pu \, dx + \int \frac{\alpha}{2} u^2 dx$$

$$\frac{\partial L}{\partial y}(y) h_y = \int (y - y_d) h_y - \nabla p \nabla(h_y) \, dx = 0$$

$$\frac{\partial L}{\partial u}(u) h_u = \int (\alpha u - p) h_u \, dx = 0$$

$$\frac{\partial L}{\partial p}(\cdot) h_p = \int -\nabla h_p \nabla y - u h_p \, dx = 0$$

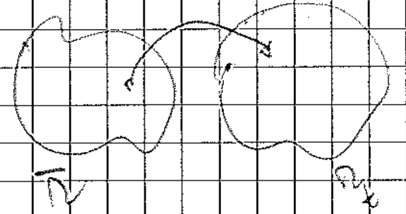
This leads to a coupled system of PDEs.

$-\Delta y = u, \quad -\Delta p = y - y_d, \quad \alpha u - p = 0$

Which again can be solved in an iterative way: $u \to y \to p$

## Shape Optimization

Consider the general pb $\min_{\Omega \in K} J(\Omega)$ where $K$ is a suitable class of compact subsets in $\mathbb{R}^d$ with regular boundary. There are two methods: topological and shape optimization. In the topological case also holes etc could be obtained. We consider shape derivatives.



Let $V: \mathbb{R}^d \to \mathbb{R}^d$ a velocity field given by $\dot{x} = V(x(t))$ and consider $\bar{x} \in \bar{\Omega}$ and $\Omega(t) = \{ x(t) : x(0) = \bar{x} \in \bar{\Omega} \}$.

Then we can define $\tilde{J}(t) := J(\Omega(t))$ and differentiate as $dJ(\bar{\Omega}, V) = \lim_{t \to 0} \frac{d}{dt} J(\Omega(t)) \Big|_{t=0}$

This is a nonlinear differential since the mapping defined by the ODE is nonlinear.

The differential is typically computed as follows:

Let $g: \mathbb{R}^d \to \mathbb{R}$ in $C^1$. ODE generates a bijection $x(t)$

$J(\Omega) = \int_\Omega g(x)\,dx$ $\quad J(\Omega(t)) = \int_{\Omega(t)} g(x)\,dx = \int_\Omega g(x(t)) \left|\det \frac{\partial x(t)}{\partial x_0}\right| dx(t)$

where $\dot{x}(t) = V(x(t))$, $x(0) = x_0$. Hence,

$$\frac{dJ(\Omega(t))}{\partial t} = \int_\Omega \nabla g(x(t)) \cdot \underbrace{V(x(t))}_{\dot x} \underbrace{\left|\det \frac{\partial x(t)}{\partial x_0}\right|}_{|M_t|} dx(t) + \int_\Omega g(x(t)) \frac{\partial M_t}{\partial t} \frac{M_t}{|M_t|}\, dx(t)$$

$\frac{\partial |z|}{\partial z} = \mathrm{sgn}(z) = \frac{z}{|z|}$

We require the derivative at $t=0$:

Hence, $x(t) = x_0 + \int_0^t V(x(s))\,ds \Rightarrow \left.\frac{\partial x(t)}{\partial x_0}\right|_{t=0} = Id. \Rightarrow M_0 = 1$

and for the derivative of $M_t$ we use Leibniz expansion:

$$M_t = \sum_{(i_1,\dots,i_d)\in \Pi_d} (-1)^{i_1+\dots+i_d} \prod_{k=1}^d \frac{(\partial x_t)_k}{(\partial x_0)_{i_k}}$$

$$\frac{\partial}{\partial t} M_t = \sum_{(\cdot)} (-1)^{\sum i_k} \sum_{j=1}^d \frac{(\partial^2 x_t)_j}{\partial (x_0)_{i_j} \partial t} \prod_{k \neq j} \frac{(\partial x_t)_k}{(\partial x_0)_{i_k}}$$

$\dot{x}_j = V_j(X(t))$

$\frac{\partial}{\partial t} \frac{\partial x_j}{\partial x_{0l}} = \nabla V_j \frac{\partial x}{\partial (x_0)_l}$

$$= \sum_{(\cdot)} (-1)^{\sum i_k} \sum_{j=1}^d \nabla V_j(x_t) \underbrace{\frac{\partial x_t}{(\partial x_0)_{i_j}}}_{(0,\dots,1,\dots,0) \text{ for } (t=0) \text{ at } i_j} \prod_{k \neq j} \frac{(\partial x_t)_k}{(\partial x_0)_{i_k}}$$

$$\overset{t=0}{=} \sum_{(\cdot)} (-1)^{\sum i_k} \sum_{j=1}^d \nabla V_j\, \delta_{j, i_j} \prod_{k \neq j} \delta_{k, i_k}$$

$\left(\frac{\partial x_t}{\partial (x_0)_{i_j}}\right)_k = \frac{\partial x_k}{\partial (x_0)_{i_j}}$

$$= \sum_{(\cdot)} (-1)^{\sum i_k} \sum_{j=1}^d \frac{\partial V_j}{\partial x_{i_j}} \prod_{k \neq j} \delta_{k, i_k} = \sum_j \frac{\partial V_j}{\partial x_j} = \mathrm{div}(V)$$

$\neq 0$ only if $k = i_k \Rightarrow j = i_j \to$ no

Leibniz: $\det A = \sum_{\tau \in S_n} \mathrm{sgn}(\tau) \prod_i^n a_{i,\tau(i)}$ permutation

$$\Rightarrow \frac{d}{dt} J(\Omega(t))\Big|_{t=0} = \int_{\Omega(0)} \nabla g(x_0) V(x_0) + g(x_0)\,\mathrm{div}\, V(x_0)\, dx_0$$

$$= \int_{\Omega} \nabla \cdot (g V)(x)\, dx \underset{\text{Gauß}}{=} \int_{\partial\Omega} g V \cdot n\, ds .$$

$\Rightarrow$ Shape derivs depend only on the boundary action of $V$!

<u>Extension to conductive case</u>

$$\min J(\Omega) \quad \text{s.t.} \quad -\nabla \cdot (a \nabla y) = 0 \quad \text{in } D$$

$$a = \begin{cases} a_1 & x \in \Omega \\ a_2 & x \in D \setminus \Omega \end{cases}$$

We denote by $e(y,\Omega) = 0$ the PDE and use again a parametrization of $\Omega_t$ as before. Then, we have that $y(t)$ is the sol. to $e(y(t), \Omega(t)) = 0$ and

$$0 = \frac{\partial e}{\partial y}(y(t), \Omega(t))\, y'(t) + \frac{\partial e}{\partial \Omega}(y(t), \Omega(t))\, V$$

In the example we have $e(y,\Omega) = 0 = \int_D a_2 \nabla y \nabla v + \int_\Omega (a_1 - a_2) \nabla y \nabla v\, dx$

Hence,

$$\frac{\partial e}{\partial y}(y,\Omega)\, y' = \int_D a \nabla y' \nabla v + \int_\Omega (a_1 - a_2) \nabla y' \nabla v\, dx = \int_D a \nabla y' \nabla v$$

and

$$\frac{\partial e}{\partial \Omega}(y,\Omega) V \underset{\text{as before}}{=} \int_{\partial\Omega} \left( (a_1 - a_2) \nabla y \nabla v \right) V \cdot n\, ds$$

Numerical approach:

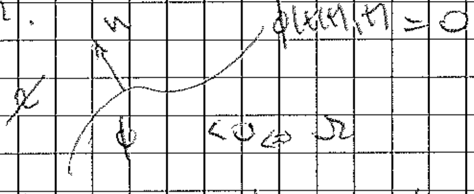
The idea is to parametrize the evolution of the boundary: $\Omega(t) = \{\phi(\cdot,t) < 0\}$

where $\phi: \mathbb{R}^d \times \mathbb{R}^+ \to \mathbb{R}$ is a differentiable fct, eg a smoothed signed distance fct $\phi(x,t) = \mathrm{dist}(x, \Omega(t))$

We choose $\phi$ st $\partial\Omega(t) = \{\phi(x,t) = 0\}$.

If now $\dot{X}(t) = V(X(t))$, then we obtain for $X(t) \in \partial\Omega(t)$ that $0 = \partial_t \phi + \nabla_x \phi \cdot \dot{X} = \partial_t \phi + (\nabla_x \phi) \cdot V$, ie. the evolution of $\phi$ is captured by a PDE.

We have seen that usually $V \cdot n$ is required to evaluated the shape derivative. Here, $n$ is the normal on $\partial\Omega$.




Since $\nabla_x \phi$ is orthogonal on its level curve and since $n$ points towards $\phi > 0$ we have

$$n(x,t) = \frac{\nabla_x \phi(x,t)}{|\nabla_x \phi(x,t)|} \qquad \forall x \in \partial\Omega(t).$$

~~Rewriting~~ ~~$V(x) =$ then~~ Since we are only interested in variations $V$ in normal direction we write $V = \overset{\in\mathbb{R}}{\tilde{V}_n} \cdot n$ and substitute this back for $\phi$:

$$0 = \partial_t \phi + \nabla_x \phi \cdot V_n \cdot n$$

$$0 = \partial_t \phi + \frac{\nabla_x \phi \cdot \nabla_x \phi}{|\nabla_x \phi|} V_n = \partial_t \phi + |\nabla \phi| V_n$$

Hence for $J(\Omega) = \int_\Omega g(x)\,dx$ we obtain

$$\frac{\partial J}{\partial \Omega}(V_n n) = \int_{\partial\Omega} g V_n \, ds$$

and the following method to update the shape $\Omega$:

$\partial\Omega^{k+1} = \{\phi(x, t_k) = 0\}$

$\partial_t \phi + V_n |\nabla_x \phi| = 0 \quad$ on $(t_k, t_k + \Delta t)$

and $\partial\Omega^{k+1} = \{\phi(x, t_{k+1}) = 0\}$.

The question is now how to choose $V_n$: Clearly we want min $J(\Omega)$ and hence $V_n$ s.t. $\frac{\partial J}{\partial V}(V_n \cdot n) < 0$

This is clearly achieved eg for $\boxed{V_n = -g}$.